

Data Visualization (CIS 568/DSC 530)

Amir Akhavan

Lecture 5: Data Types

Monday September 14, 2020

Datasets

➔ Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

➔ Data and Dataset Types

Tables

Items

Attributes

Networks &
Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

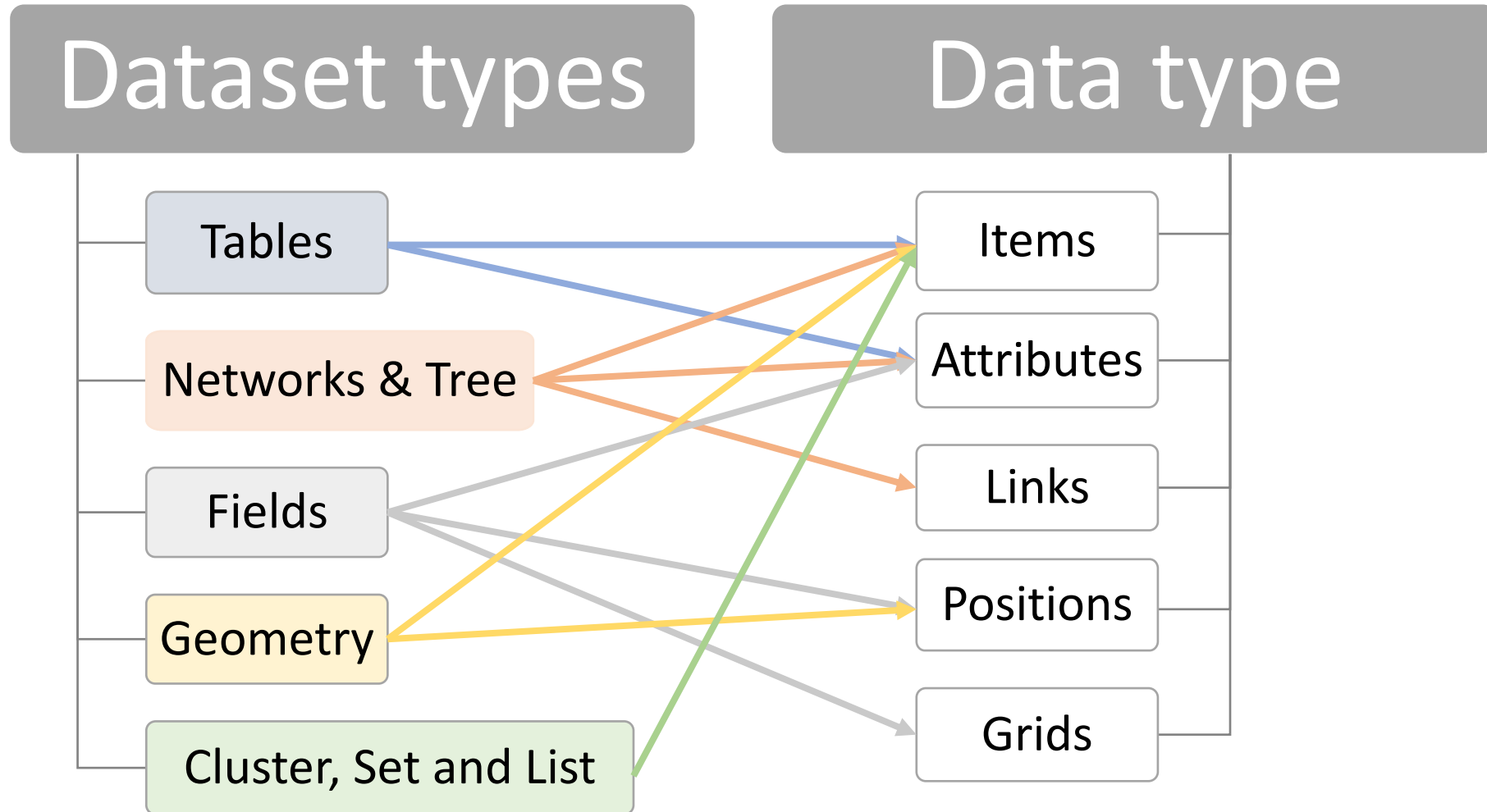
Items

Positions

Clusters,
Sets, Lists

Items

Data Types



Attributes

➔ Attribute Types

➔ Categorical



➔ Ordered

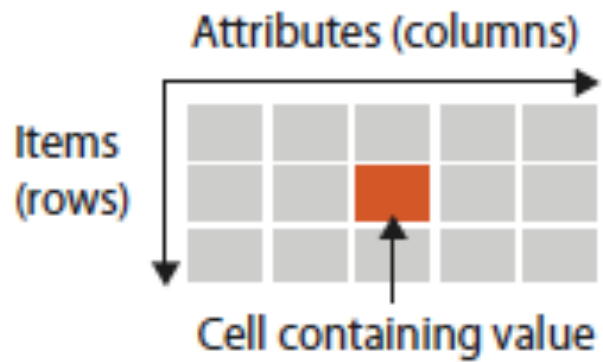
➔ *Ordinal*



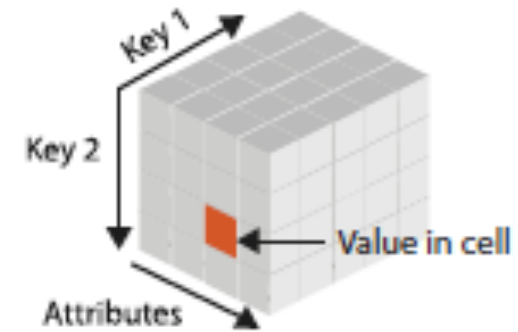
➔ *Quantitative*



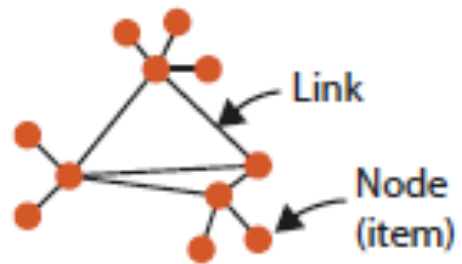
→ Tables



→ *Multidimensional Table*



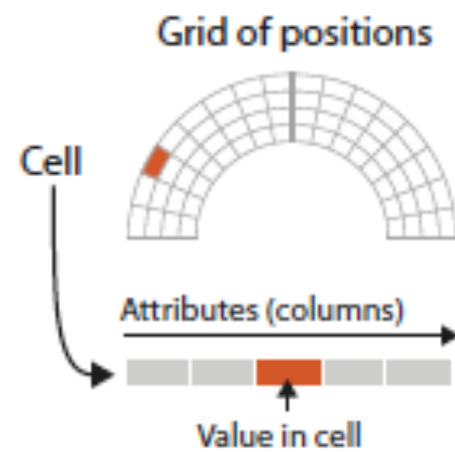
→ Networks



→ *Trees*



→ Fields (Continuous)



→ Geometry (Spatial)



➔ Dataset Availability

➔ Static



Census data
WHO
UNHCR

➔ Dynamic



Live twitter data
Earthquake data
NASA Merra 2

Semantics

```
graph TD; S[Semantics] --- S1[real-world meaning of Data]; T[Type] --- T1[structural or mathematical interpretation of Data]
```

real-world
meaning of Data

Type

structural or
mathematical
interpretation of Data

| ID | Name | Age | Shirt Size | Favorite Fruit |
|----|---------|-----|------------|----------------|
| 1 | Amy | 8 | S | Apple |
| 2 | Basil | 7 | S | Pear |
| 3 | Clara | 9 | M | Durian |
| 4 | Desmond | 13 | L | Elderberry |
| 5 | Ernest | 12 | L | Peach |
| 6 | Fanny | 10 | S | Lychee |
| 7 | George | 9 | M | Orange |
| 8 | Hector | 8 | L | Loquat |
| 9 | Ida | 10 | M | Pear |
| 10 | Amy | 12 | M | Orange |

Attributes

➔ Attribute Types

➔ Categorical



➔ Ordered

➔ Ordinal

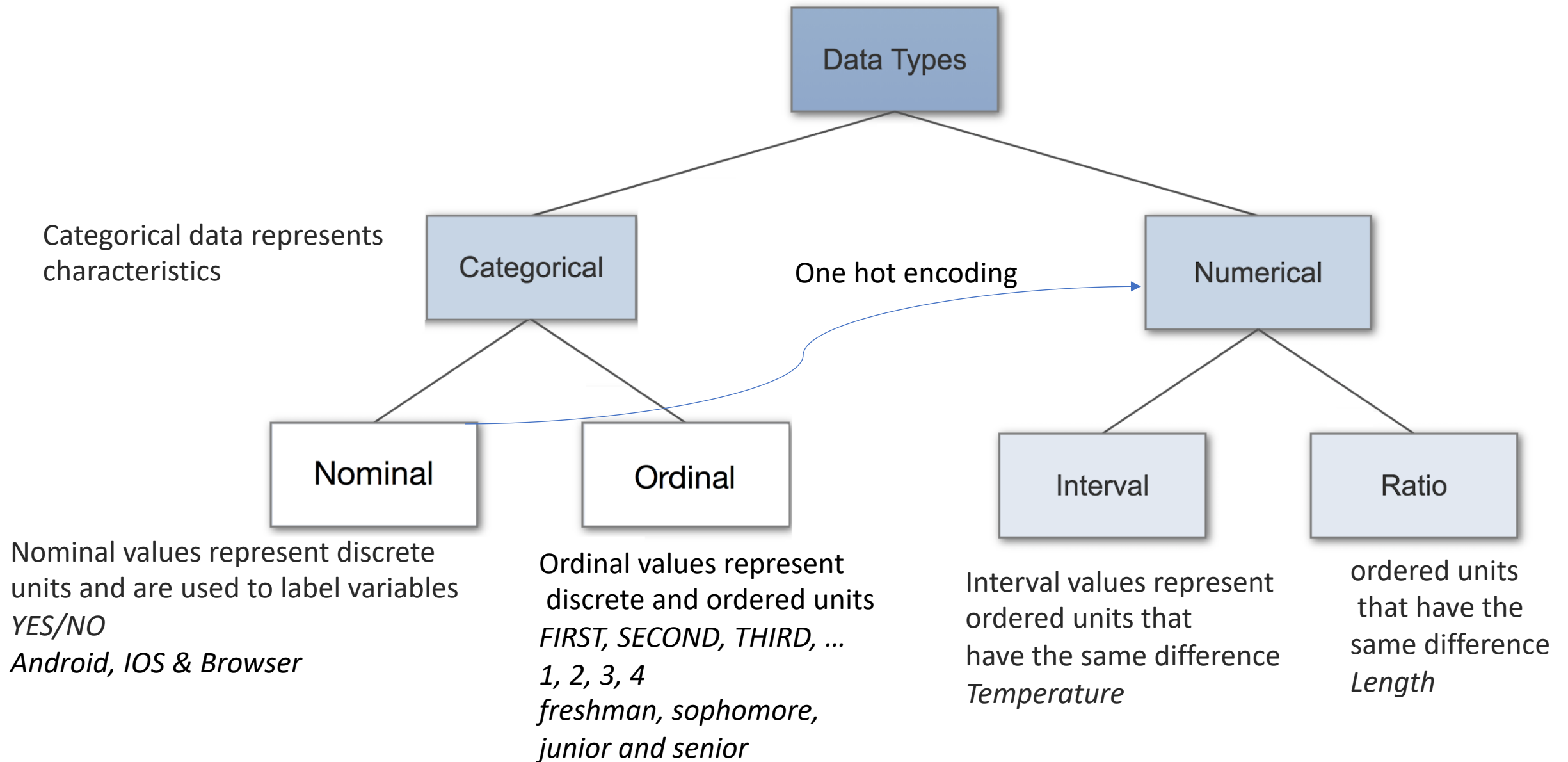


➔ Quantitative



| A | B | C | S | T | U |
|----------|------------|-----------------|-------------------|---------------------|-----------|
| Order ID | Order Date | Order Priority | Product Container | Product Base Margin | Ship Date |
| 3 | 10/14/06 | 5-Low | Large Box | 0.8 | 10/21/06 |
| 6 | 2/21/08 | 4-Not Specified | Small Pack | 0.55 | 2/22/08 |
| 32 | 7/16/07 | 2-High | Small Pack | 0.79 | 7/17/07 |
| 32 | 7/16/07 | 2-High | Jumbo Box | 0.72 | 7/17/07 |
| 32 | 7/16/07 | 2-High | Medium Box | 0.6 | 7/18/07 |
| 32 | 7/16/07 | 2-High | Medium Box | 0.65 | 7/18/07 |
| 35 | 10/23/07 | 4-Not Specified | Wrap Bag | 0.52 | 10/24/07 |
| 35 | 10/23/07 | 4-Not Specified | Small Box | 0.58 | 10/25/07 |
| 36 | 11/3/07 | 1-Urgent | Small Box | 0.55 | 11/3/07 |
| 65 | 3/18/07 | 1-Urgent | Small Pack | 0.49 | 3/19/07 |
| 66 | 1/20/05 | 5-Low | Wrap Bag | 0.56 | 1/20/05 |
| 69 | 6/4/05 | 4-Not Specified | Small Pack | 0.44 | 6/6/05 |
| 69 | 6/4/05 | 4-Not Specified | Small Pack | 0.6 | 6/6/05 |
| 70 | 12/18/06 | 5-Low | Small Pack | 0.59 | 12/23/06 |
| 70 | 12/18/06 | 5-Low | Small Pack | 0.82 | 12/23/06 |
| 96 | 4/17/05 | 2-High | Small Pack | 0.55 | 4/19/05 |
| 97 | 1/29/06 | 3-Medium | Small Pack | 0.38 | 1/30/06 |
| 129 | 11/19/08 | 5-Low | Small Pack | 0.37 | 11/28/08 |
| 130 | 5/8/08 | 2-High | Small Box | 0.37 | 5/9/08 |
| 130 | 5/8/08 | 2-High | Medium Box | 0.38 | 5/10/08 |
| 130 | 5/8/08 | 2-High | Small Box | 0.6 | 5/11/08 |
| 132 | 6/11/06 | 3-Medium | Medium Box | 0.6 | 6/12/06 |
| 132 | 6/11/06 | 3-Medium | Jumbo Box | 0.69 | 6/14/06 |
| 134 | 5/1/08 | 4-Not Specified | Large Box | 0.82 | 5/3/08 |
| 135 | 10/21/07 | 4-Not Specified | Small Pack | 0.64 | 10/23/07 |
| 166 | 9/12/07 | 2-High | Small Box | 0.55 | 9/14/07 |
| 193 | 8/8/06 | 1-Urgent | Medium Box | 0.57 | 8/10/06 |
| 194 | 4/5/08 | 3-Medium | Wrap Bag | 0.42 | 4/7/08 |

quantitative
ordinal
categorical



Transforming Nominal data into a Numeric type

One Hot Encoding

| Color | | Red | Yellow | Green |
|--------|--|-----|--------|-------|
| Red | | 1 | 0 | 0 |
| Red | | 1 | 0 | 0 |
| Yellow | | 0 | 1 | 0 |
| Green | | 0 | 0 | 1 |
| Yellow | | 0 | 0 | 1 |

<https://www.kaggle.com/dansbecker/using-categorical-data-with-one-hot-encoding>

Key versus Value Semantics

A key attribute acts as an index that is used to look up value attributes.

- Flat Tables
- Multidimensional Tables
- Fields
 - Scalar Fields
 - Vector Fields
 - Tensor Fields
 - Field Semantics

In this table

Item: row

Attribute: column

Value: intersection of row & column

| A | B | C | S | T | U |
|----------|------------|-----------------|-------------------|---------------------|-----------|
| Order ID | Order Date | Order Priority | Product Container | Product Base Margin | Ship Date |
| 3 | 10/14/06 | 5-Low | Large Box | 0.8 | 10/21/06 |
| 6 | 2/21/08 | 4-Not Specified | Small Pack | 0.55 | 2/22/08 |
| 32 | 7/16/07 | 2-High | Small Pack | 0.79 | 7/17/07 |
| 32 | 7/16/07 | 2-High | Jumbo Box | | 7/17/07 |
| 32 | 7/16/07 | 2-High | Medium Box | | 7/18/07 |
| 32 | 7/16/07 | 2-High | Medium Box | | 7/18/07 |
| 35 | 10/23/07 | 4-Not Specified | Wrap Bag | 0.52 | 10/24/07 |
| 35 | 10/23/07 | 4-Not Specified | Small Box | 0.58 | 10/25/07 |
| 36 | 11/3/07 | 1-Urgent | Small Box | 0.55 | 11/3/07 |
| 65 | 3/18/07 | 1-Urgent | Small Pack | 0.49 | 3/19/07 |
| 66 | 1/20/05 | 5-Low | Wrap Bag | 0.56 | 1/20/05 |
| 69 | 5/4/05 | 4-Not Specified | Small Pack | 0.44 | 6/6/05 |
| 69 | 5/4/05 | 4-Not Specified | Wrap Bag | 0.6 | 6/6/05 |
| 70 | 12/18/06 | 5-Low | Small Box | 0.59 | 12/23/06 |
| 70 | 12/18/06 | 5-Low | Wrap Bag | 0.82 | 12/23/06 |
| 96 | 4/17/05 | 2-High | Small Box | 0.55 | 4/19/05 |
| 97 | 1/29/06 | 3-Medium | Small Box | 0.38 | 1/30/06 |
| 129 | 11/19/08 | 5-Low | Small Box | 0.37 | 11/28/08 |
| 130 | 5/8/08 | 2-High | Small Box | 0.37 | 5/9/08 |
| 130 | 5/8/08 | 2-High | Medium Box | 0.38 | 5/10/08 |
| 130 | 5/8/08 | 2-High | Small Box | 0.6 | 5/11/08 |
| 132 | 6/11/06 | 3-Medium | Medium Box | 0.6 | 6/12/06 |
| 132 | 6/11/06 | 3-Medium | Jumbo Box | 0.69 | 6/14/06 |
| 134 | 5/1/08 | 4-Not Specified | Large Box | 0.82 | 5/3/08 |
| 135 | 10/21/07 | 4-Not Specified | Small Pack | 0.64 | 10/23/07 |
| 166 | 9/12/07 | 2-High | Small Box | 0.55 | 9/14/07 |
| 193 | 8/8/06 | 1-Urgent | Medium Box | 0.57 | 8/10/06 |
| 194 | 4/5/08 | 3-Medium | Wrap Bag | 0.42 | 4/7/08 |

attribute

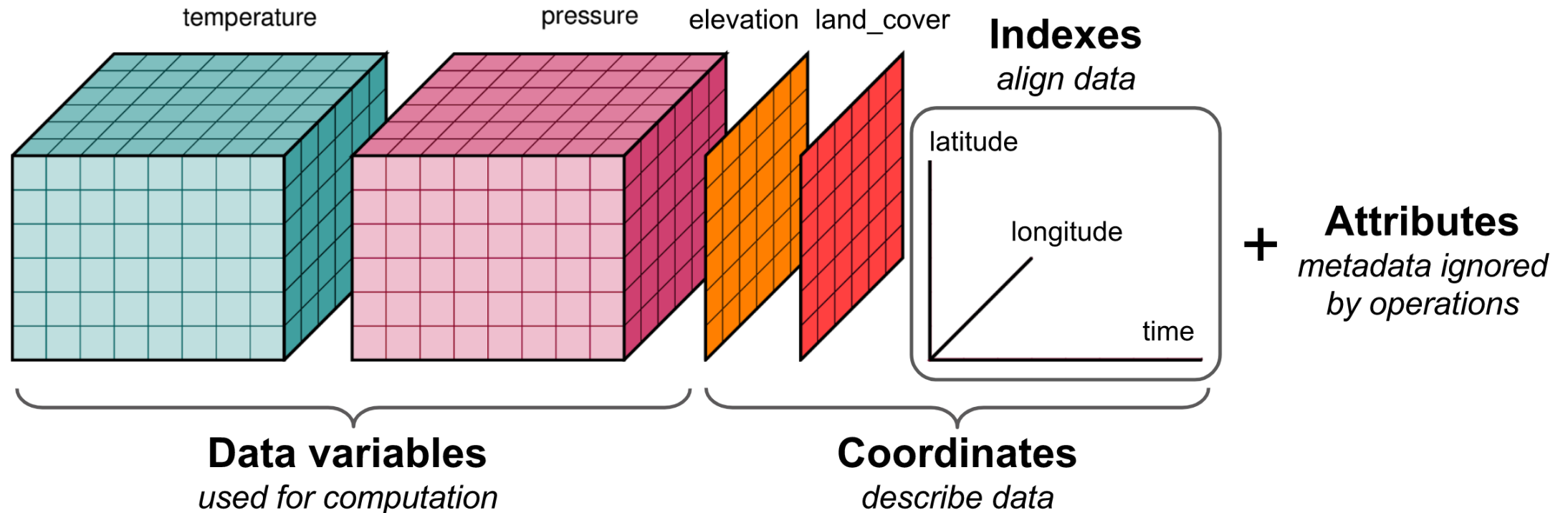
item

cell

Fields

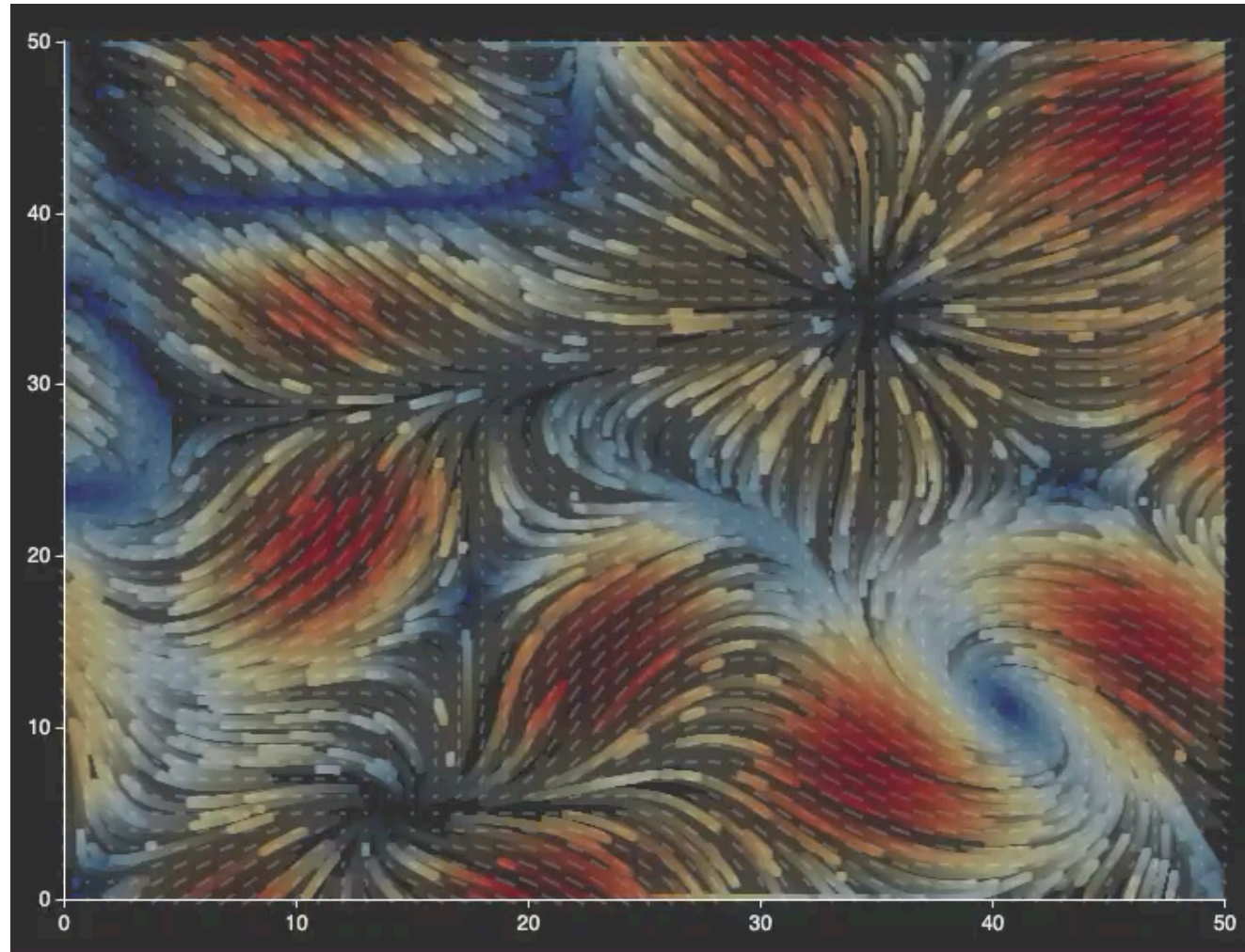
- Fields like tables are characterized by “Key” vs “Values”
- Fields represent “continuous” values
- Fields’ Multivariate structure depends on the number of value attributes
- Fields’ Multidimensional structure depends on the number of keys

Scalar Fields: Multidimensional & Multivariate

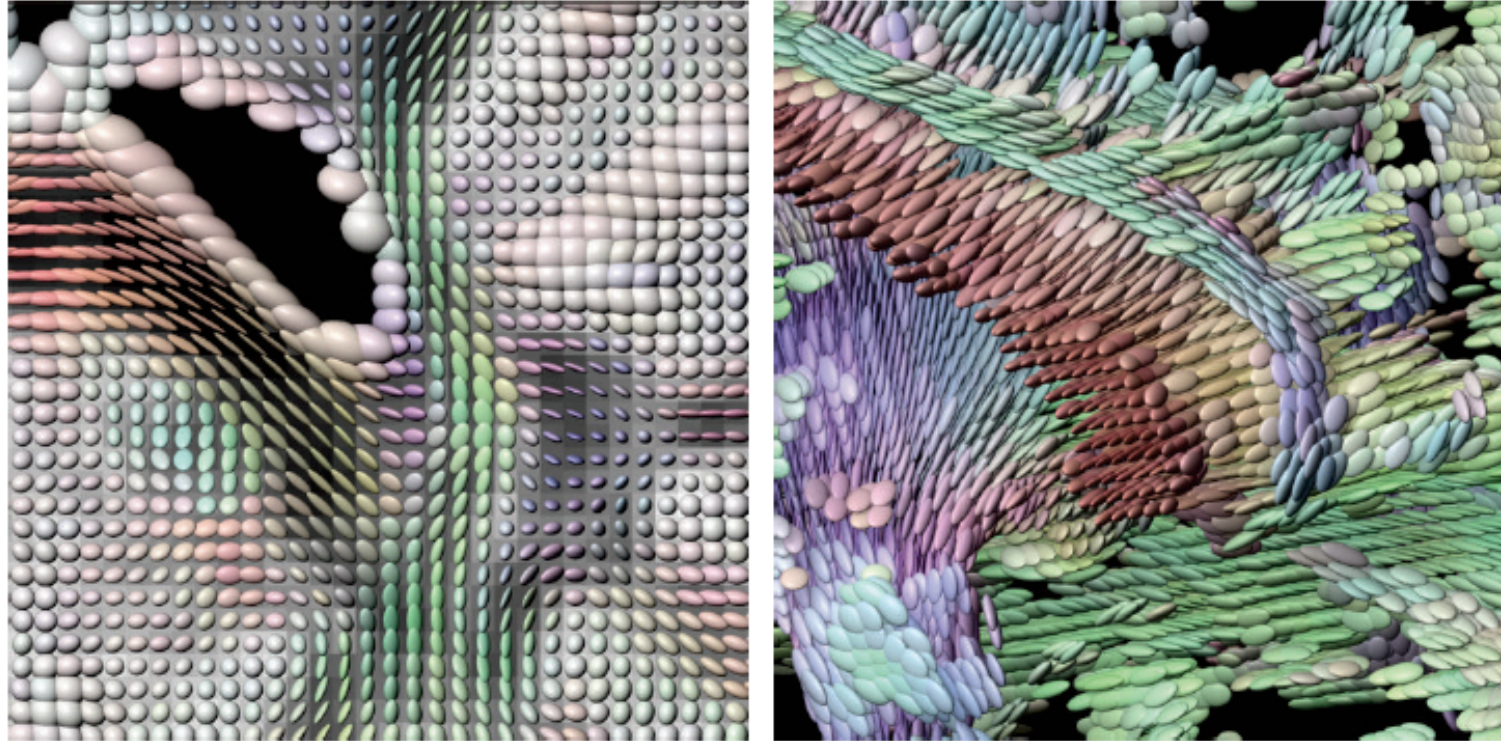


<http://xarray.pydata.org/en/stable/dask.html>

Vector Field



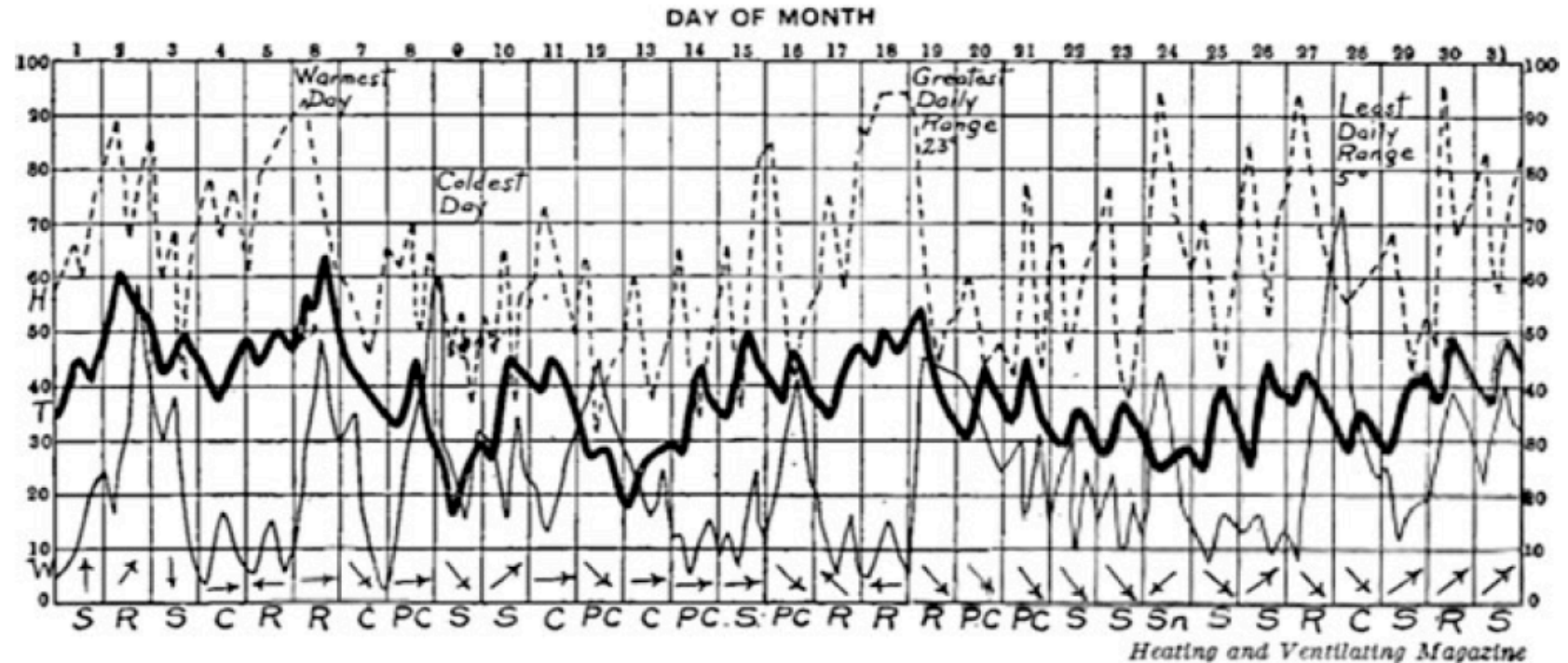
Tensor Field



Ellipsoid glyphs show shape and orientation of tensors at each cell in a field. (a) 2D slice. (b) 3D field, with isotropic glyphs filtered out. From [Kindlmann04, Figures 10a and 11a].

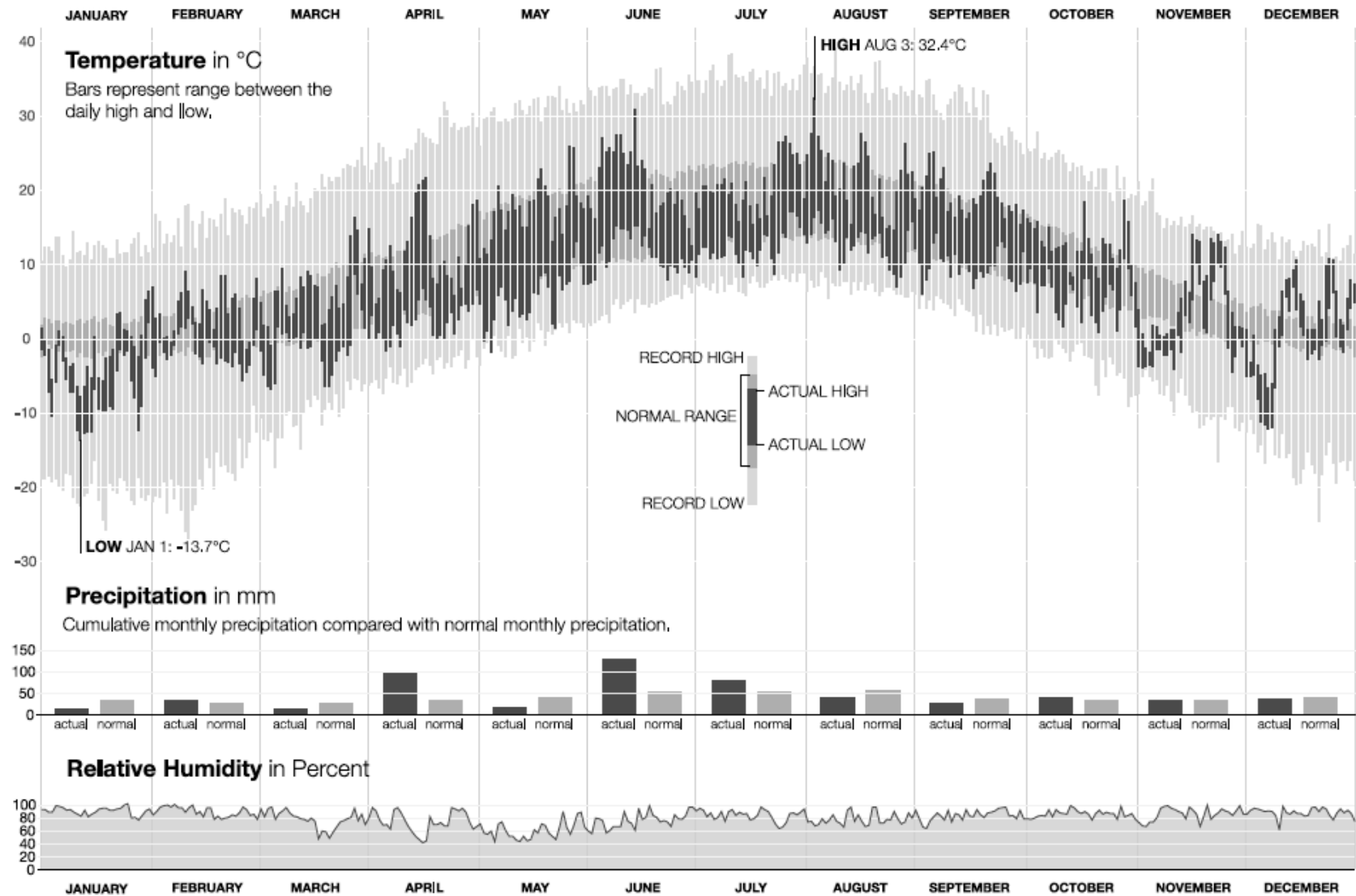
Temporal Semantics (Time-varying Data)

New York City weather 12/1912.
dark-line: Temperature
Light-line: Wind velocity
dotted line: relative humidity (%)



Weather in 1980

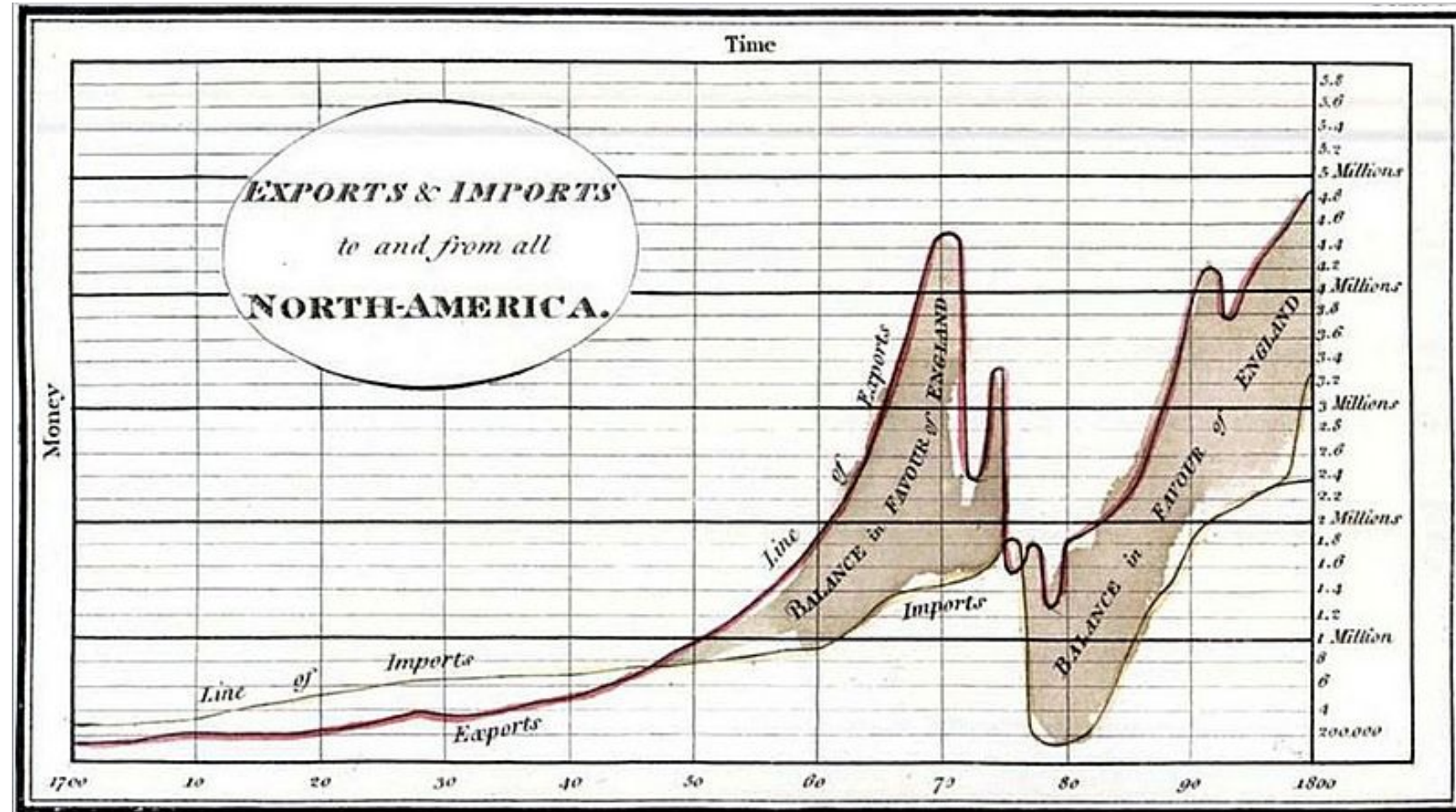
New York City weather 1980



Temporal Charts (William Playfair)

Export and Import to and from all
North America from 1700 to 1800
Playfair (1801)

dark-line: Exports
light-line: Imports



Reading

- <http://ieg.ifs.tuwien.ac.at/~aigner/timeviz-book/> (Chapter 2)
- Munzner' Book (Chapter 2)
- <http://xarray.pydata.org/en/stable/dask.html>